The capsule and genetic background, rather than specific loci, strongly influence pneumococcal growth kinetics

Chrispin Chaguza[1,2,3,4‡], Daan W. Arends[5], Stephanie W. Lo[3], Indri Hapsari Putri[5], Anna York[1], John A. Lees[6], Anne L. Wyllie[1], Daniel M. Weinberger[1], Stephen D. Bentley[3], Marien I. de Jonge[5#], Amelieke J.H. Cremers[5,7‡#]

[1]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, Connecticut, USA

[2]Yale Institute for Global Health, Yale University, New Haven, Connecticut, USA

[3]Parasites and Microbes Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, United Kingdom

[4]Department of Host-Microbe Interactions, St Jude Children's Research Hospital, Memphis, Tennessee, USA

[5]Laboratory of Medical Immunology, Radboud Institute of Molecular Life Sciences, Radboudumc Center for Infectious Diseases, Radboudumc, Nijmegen, The Netherlands

[6]European Molecular Biology Laboratory–European Bioinformatics Institute, Cambridge, United Kingdom

[7]Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland

#Senior author

*These authors contributed equally

‡Correspondence: Chrispin.Chaguza@yale.edu (C.C.) and Amelieke.Cremers@radboudumc.nl (A.J.H.C.)

Abstract

Bacterial growth characteristics intrinsic to each strain can impact and influence gene expression, antibiotic susceptibility, and disease pathogenesis. However, little is known about specific genomic variations that influence these bacterial growth features. Here, we investigate the impact of *Streptococcus pneumoniae* genetics on its *in vitr*o growth features to shed light on genes that may be important targets in the development of vaccines and therapeutics. We use statistical models to estimate growth features and show that they varied significantly across capsular serotypes and lineages, strongly correlated with the phylogeny, and were highly heritable, highlighting a strong genetic basis. Despite this, genome-wide association studies revealed no specific genomic loci statistically associated with the growth features independently of the genetic background, including those in the capsular polysaccharide synthesis locus. Our findings show that the serotype and lineage, and possibly a combination of genomic loci, influence intrinsic pneumococcal growth kinetics, which may have implications on pneumococcal disease pathogenesis.

Keywords

Streptococcus pneumoniae; *in vitro* growth; GWAS; population genomics; capsular serotype; genetic background

45 # Introduction

46 Once termed "the captain of the men of death" by Sir William Osler in the 20th century,
47 *Streptococcus pneumoniae*, or the pneumococcus, remains a significant cause of life-threatening
48 invasive diseases, such as sepsis, pneumonia, and meningitis[1], despite the widespread
49 implementation of life-saving serotype-specific polysaccharide pneumococcal conjugate
50 vaccines (PCVs)[2]. The high capsular diversity, amounting to ~100 serotypes with varying
51 potential for colonising the nasopharynx and causing invasive diseases[3,4] and mortality[5–7],
52 demonstrates the remarkable genetic diversity of the pneumococcus. Such high genetic diversity
53 reflects the plasticity of the pneumococcal genome, which readily swaps genetic material with
54 other strains through recombination and horizontal gene transfer[8]. However, the impact of
55 pneumococcal genomic variation on its phenotypes, especially those associated with disease
56 pathogenesis and clinical manifestations, remains less understood.
57
58 The pneumococcus possesses an arsenal of virulence factors. The extracellular polysaccharide
59 capsule is the main virulence factor critical for transmissibility, survival, and pathogenicity in the
60 human host[9–11]. Recent studies in the pneumococcus and other bacterial pathogens suggest that
61 variability in growth kinetics impacts disease pathogenesis, clinical manifestation, and antibiotic
62 therapy. For example, differential pneumococcal growth features appear to correlate with fitness
63 for colonisation and invasive disease,[12] clinical manifestation[12–14] and response to environmental
64 conditions[14–16]. Studies of other bacterial pathogens suggest that growth rate influences gene and
65 protein expression[17,18], adhesion[19], bacterial competition[20], plasmid replication[21], bacteriophage
66 dynamics[22], and antibiotic accumulation and efficacy[23–25]. Similarly, experimental and
67 mathematical modelling studies have demonstrated that bacterial growth correlates with bacterial
68 load and host-mediated cell lysis and impacts dispersal to different host cells and tissues during
69 systemic infection[26]. Previous studies of pneumococcus and other bacteria have suggested the
70 influence of environmental factors[15,16] and bacterial-specific factors, including capsular
71 serotype[14,27] and the presence of plasmids[28] on bacterial growth phenotypes. These growth
72 phenotypes may consequently impact invasive pneumococcal disease pathogenesis[12–14].

However, the specific pneumococcal genomic variations affecting growth kinetics remain less understood.

Studying bacterial growth kinetics sheds light on the influence of genes, genomic variation, and growth media conditions on bacterial fitness[29,30]. However, studying bacterial growth kinetics *in vivo* remains intrinsically challenging due to the complexity of the conditions in the hosts, although innovative approaches are being developed[31,32]. On the other hand, *in vitro* methods provide a valuable alternative for assessing the impact of specific bacterial and environmental factors on bacterial fitness. Although these *in vitro* methods have revealed remarkable insights on bacterial fitness, these experiments have mostly yielded strain-specific rather than population-level insights due to the analysis of a small number of isolates from few genetic backgrounds.

Here, we studied the impact of pneumococcal genetics on *in vitro* growth kinetics to understand the influence of bacterial genetic variation on pneumococcal pathogenesis. Specifically, we performed high-throughput quantification of the *in vitro* growth kinetics and whole genome sequencing of 348 invasive pneumococcal isolates from the Netherlands to determine the impact of genomic variation on growth kinetics. We conducted growth curve experiments on 348 isolates in blood-like media conditions, and absorbance was monitored over 15 hours. Growth features were determined from these curves using statistical models, and we assessed the impact of capsular serotype and genetic background on the estimated growth features and tested their association with the phylogeny of the isolates. In addition, we used a genome-wide association study (GWAS) approach to identify specific genomic loci associated with the growth features independent of the genetic background. Finally, we calculated the variability in the growth features explained by pneumococcal genetics. Our study suggests that the capsular serotype, genetic background or lineage, and a combination of genetic loci may influence pneumococcal

growth kinetics, a fundamental trait potentially influencing nasopharyngeal colonisation, invasive disease pathogenesis, and clinical manifestations.

Results

Population structure of a diverse collection of Dutch invasive pneumococcal isolates and derivation of their growth kinetics

We aimed to understand the impact of the capsular serotype, genetic background, and specific genomic loci, independent of the clonal population structure, on the pneumococcal growth features **(Fig. 1 and Supplementary Data 1)**. Therefore, we investigated the *in vitro* growth kinetics of 348 pneumococcal isolates collected from patients with invasive pneumococcal diseases via the Pneumococcal Bacteraemia Collection Nijmegen (PBCN) cohort study at two hospitals in Nijmegen, the Netherlands before (2000-2006) and after (2007-2011) the introduction of the seven-valent pneumococcal conjugate vaccine (PCV) [PCV7] into the Dutch paediatric immunisation programme[33]. Our isolates represent 34 capsular serotypes, 40 pneumococcal lineages, defined based on the global pneumococcal sequence cluster (GPSC) nomenclature[34], and 99 sequence types (STs) defined by the pneumococcal multilocus sequence typing (MLST) scheme[35]. An interactive and annotated maximum likelihood phylogeny of the isolates used in this study is available online at the Microreact platform for open data visualisation and sharing for genomic epidemiology (https://microreact.org/project/wxmdblfgprepwugvbbtaq7). We estimated a Simpson diversity index of 0.932 (95% confidence interval [CI]: 0.924 to 0.940), 0.937 (95% CI: 0.928 to 0.945), and 0.958 (95% CI: 0.949 to 0.966) for the serotypes, lineages, and STs, respectively. The high richness and diversity of pneumococcal serotypes and lineages highlighted a substantial population-level diversity of the study isolates in the Netherlands.

Capsular serotype and genetic background significantly impact pneumococcal growth kinetics

To explore the impact of capsular serotype and genetic background or lineage on the pneumococcal growth features estimated from a fitted growth curve, namely lag-phase duration,

maximum growth density (H_{max}), maximum growth change (ΔH), and average growth rate (r) **(Fig. 1, 2, and Supplementary Data 1),** we assessed the relationship between these growth features and serotypes and lineages associated with the study isolates **(Fig. 3a)**. In general, we found significant variability of the growth features across serotypes: average growth rate, maximum growth density, maximum growth change, and lag phase duration ($P<0.0001$; Kruskal-Wallis test) consistent with previous studies[14,15] **(Fig. 3b-e and Supplementary Data 2)**. We observed similar patterns across different pneumococcal lineages **(Fig. 3f-i and Supplementary Data 2)**.

Recombination and horizontal gene transfer may lead to the replacement of part or the entire capsule biosynthesis locus resulting in the production of a polysaccharide capsule distinct from the original one, a phenomenon called capsule-switching[36]. Consequently, some lineages are associated with multiple serotypes, and similarly, a single serotype can be expressed in distinct genetic backgrounds[37]. Examples of lineages containing multiple serotypes included GPSC3, GPSC4, and GPSC7, and likewise, serotypes, including 3, 19A, and 19F, were expressed in multiple genetic backgrounds, which also express other serotypes **(Fig. 3a)**. Overall, the average growth rate for serotypes expressed in different genetic backgrounds broadly showed similar patterns. For example, lineages expressing serotype 14 had consistently higher average growth rates. In contrast, lineages expressing serotype 3 showed lower rates, consistent with the notion that capsule production interferes with pneumococcal growth[14] **(Fig. 3b).** Therefore, we investigated the impact of the capsular serotype and genetic background on the estimated pneumococcal growth features.

First, we identified lineages that expressed multiple serotypes and serotypes found across multiple lineages. To ensure robust statistical comparison, we only focused on serotype-lineage combinations with at least three sequenced isolates in our study. We found two lineages, with enough sequenced isolates for comparison, GPSC3 and GPSC7, which expressed multiple

152 serotypes **(Fig. 3j and Supplementary Data 2)**. We noted higher growth rates for serotype 33F
153 than serotype 8 (*P*=0.0007; Kruskal-Wallis test).
154
155 In contrast, we found no statistically significant differences in the growth rates between
156 serotypes 23A and 23F in lineage GPSC7 (*P*=0.2568; Kruskal-Wallis test). We speculated that
157 the absence of statistically significant differences between serotypes 23A and 23F reflected their
158 presence in the same serogroup, therefore, had similar capsule properties. Interestingly, we
159 observed similar patterns for the maximum growth density, maximum growth change, and lag
160 phase duration features between serotypes on the GPSC3 and GPSC7 lineages **(Fig. 3k-m and**
161 **Supplementary Data 2)**. Next, we investigated the growth rates of the same serotypes expressed
162 in distinct genetic backgrounds. Although some serotypes showed similar growth rates across
163 distinct genetic backgrounds, such as serotype 3 (*P*=0.4599; Kruskal-Wallis test), others showed
164 different growth rates across different lineages, as seen with serotypes 14 (*P*=0.0006; Kruskal-
165 Wallis test) and 19F (*P*=0.0105; Kruskal-Wallis test). We noted similar patterns for the other
166 growth features: maximum growth density, maximum growth change, and lag phase duration
167 **(Fig. 3n-q, Supplementary Data 2)**.
168
169 These findings demonstrate that the capsular serotype and genetic background significantly
170 impact pneumococcal growth kinetics. The resultant growth features likely depend on the
171 interaction between these factors, whereby some combinations may result in more pronounced
172 growth effects than others.
173

174 Pneumococcal growth kinetics show heritability and correlation with the 175 phylogeny of the isolates

176 Based on the findings that both capsular serotype and genetic background are associated with
177 growth kinetics, we hypothesised that the estimated pneumococcal growth features exhibit a high
178 phylogenetic signal. We assessed the phylogenetic signal for each growth feature using Pagel's λ
179 statistic[38], which evaluates the correlation between given traits and the phylogenetic tree of the
180 isolates, with values close to zero and one indicating weaker and stronger phylogenetic signals,
181 respectively. By reconstructing the estimated growth features on the internal nodes of the

182 phylogeny, we found high but variable associations between the mapped features and different
183 genetic backgrounds, which suggested potentially strong phylogenetic signals **(Fig. 4,**
184 **Supplementary Fig. 2 and 3, and Supplementary Data 3)**.
185
186 The amount of the phylogenetic signals based on the normal transformed growth features was
187 highest for the average growth rate (Pagel's λ=0.72, 95% CI: 0.61 to 0.81; $P<2.08\times10^{-64}$),
188 maximum growth change (Pagel's λ=0.60, 95% CI: 0.40 to 0.76; $P<1.53\times10^{-16}$), and maximum
189 growth density (Pagel's λ=0.73, 95% CI: 0.57 to 0.84; $P<4.18\times10^{-26}$), while the lag phase
190 duration showed a lower phylogenetic signal (Pagel's λ=0.30, 95% CI: 0.17 to 0.47; $P<5.42\times10^{-}$
191 ^{16}) **(Fig. 5a)**. These findings suggest that there is a heterogeneous correlation of the growth
192 features with the phylogeny.
193
194 To quantify the variability in the pneumococcal growth kinetics explained by the genotype, we
195 calculated the narrow-sense heritability (h^2) for the four normal-transformed estimated growth
196 features and the maximum growth rate from Arends et al[13]. We found strong, though not equal,
197 evidence for a genetic basis for all the estimated growth features: average growth rate (h^2=0.76,
198 95% CI: 0.68 to 0.85), maximum growth change (h^2=0.80, 95% CI: 0.71 to 0.89), maximum
199 growth density (h^2=0.88, 95% CI: 0.83 to 0.94), and lag phase duration (h^2=0.29, 95% CI: 0.14
200 to 0.43) **(Fig. 5b, Supplementary Data 3)**. To further understand the impact of the capsular
201 serotype and genetic background, we fitted a linear mixed effects model for each growth
202 parameter with serotype and lineage (defined by GPSCs[34]) as random effects to estimate the
203 amount of variance of the growth features explained by the serotype (h^2_{serotype}) and lineage or
204 GPSC (h^2_{GPSC}) calculated as the ratio of the variance of the serotype and lineage components
205 relative to the total variance, including the residual or environmental effects. We calculated
206 h^2_{serotype}=0.47 and h^2_{GPSC}=0.11 for the average growth rate, h^2_{serotype}=0.13 and h^2_{GPSC}=0.28 for the
207 maximum growth change, h^2_{serotype}=0.29 and h^2_{GPSC}=0.46 for the maximum growth,
208 h^2_{serotype}=0.08 and h^2_{GPSC}=0.39 for the lag phase duration feature.
209
210 Altogether, these findings demonstrate a substantial impact of the serotype and lineage on the
211 variation of the pneumococcal growth features, whereby most serotypes tend to have similar
212 characteristics regardless of the genetic background. Similarly, some genetic backgrounds tend

213 to show similar phenotypic patterns irrespective of the serotypes, but the serotype appears to
214 have a more significant influence on the growth characteristics.

215 No individual genomic loci are associated with pneumococcal growth 216 features

217 Having found moderate to high narrow-sense heritability of the growth features, we then
218 undertook a GWAS analysis to identify specific genomic loci associated with these features
219 independent of the strains' genetic background using a linear mixed-effects model. We
220 performed separate GWAS analyses of the maximum growth change, maximum growth density,
221 maximum growth change, and lag phase duration phenotypes derived from the growth curve. We
222 conducted the GWAS using variable-length *k*-mer sequences or unitigs to capture genetic
223 variation in the pneumococcal genomes at different levels, including SNPs, multi-allelic sites,
224 and insertions and deletions[39,40]. We found no individual unitigs associated with the derived
225 growth feature phenotypes **(Fig. 6a, Supplementary Data 4)**. Furthermore, repeating the GWAS
226 analyses adjusting for the serotype revealed no statistically significant hits associated with the
227 derived growth phenotypes, suggesting that the variants in the pneumococcal strains are
228 correlated with the growth kinetics independent of the serotype **(Fig. 6b, Supplementary Data
229 4)**. Overall, the observed and expected statistical significance values in the QQ-plots for our
230 study revealed no apparent issues with the population structure of the isolates, which is a
231 significant confounder in similar studies[41] **(Fig. 6c, d)**. Overall, these findings suggest that no
232 individual genomic variation influences the pneumococcal growth rates, suggesting that multiple
233 genetic loci and lineage effects modulate pneumococcal growth kinetics.

234 Average *in vitro* growth rates are not strongly correlated with a serotype's 235 invasiveness or its prevalence in carriage or disease

236 Previous studies have elucidated that bacterial growth rate may influence colonisation and
237 disease[16], including systemic dissemination during infection[26] and disease manifestations[13,27].
238 Pneumococcal strains that are prolific growers may display better survival and replication in
239 systemic tissues than slow growers without vaccine-induced non-serotype-specific immunity.
240 We used the isolates collected before the introduction of the PCV7 vaccine in the Netherlands[42]
241 to assess the association between *in vitro* growth rate of the isolates and their population-level

242 characteristics: prevalence in invasive disease and nasopharyngeal carriage and invasiveness, i.e.,
243 odds of being isolated in invasive disease compared to the carriage state. Our first hypothesis
244 was that the average *in vitro* growth rate for each of the pneumococcal serotypes is positively
245 associated with the prevalence of serotypes in invasive diseases. Our analysis revealed no
246 evidence of an association between the prevalence of the serotypes in invasive disease and the *in*
247 *vitro* growth rate (Spearman correlation: -0.30, P=0.0856) **(Fig. 7a, Supplementary Data 4)**.
248 Second, we hypothesized that the average *in vitro* growth rate of the serotypes is positively
249 associated with the prevalence of the serotypes in the nasopharyngeal carriage as fitter strains
250 would grow faster than less fit strains. However, contrary to our hypothesis, we found a negative
251 association between the prevalence of the serotypes in carriage and the *in vitro* growth rate
252 (Spearman correlation=-0.53, P=0.0098) **(Fig. 7b, Supplementary Data 4)**. We repeated the
253 analysis after excluding serotype 3 and found no evidence of an association between growth rate
254 and carriage prevalence (Spearman correlation=0.38, P=0.0925) which suggests that serotype 3
255 alone was driving the previously observed association. Finally, we tested the hypothesis that the
256 growth rate is positively associated with the invasiveness of the serotypes. We found no
257 association between the invasiveness and the *in vitro* growth rates of the serotypes (Spearman
258 correlation: 0.35, P=0.1024) **(Fig. 7c and Supplementary Data 4)**. These findings suggest that
259 growth rates alone may not be sufficient to explain the prevalence and invasiveness of
260 pneumococcal serotypes, which suggests the influence of other pneumococcal, microbiome (e.g.,
261 viral infection), and host factors in determining pneumococcal disease susceptibility.

262 Discussion

263 Bacterial growth kinetics impact several traits, including gene and protein expression[17,18],
264 antibiotic accumulation and efficacy[23–25], cell surface adhesion[19], intra- and inter-specific
265 competition[20], response to environmental conditions[14–16], plasmid replication[21], bacteriophage
266 dynamics[22], colonisation and disease pathogenesis[12], including clinical manifestation[12–14]. Whilst
267 previous studies have attempted to understand bacterial, host, and environmental factors
268 influencing bacterial growth, a major drawback has been a narrow focus on a small subset of
269 strains[43]. Therefore, the insights gained from these studies have mostly been for specific strains,
270 as they ignored potential variability across different genetic backgrounds. These population-level
271 studies are critical for genetically diverse bacterial pathogens with highly structured

populations[34], substantial antigenic diversity[44], and recombination rates, such as the pneumococcus[8,36,45]. Here, we examine the association between pneumococcal genetics and growth kinetics using a combination of high-throughput quantification of *in vitro* pneumococcal growth, whole-genome sequencing, and population genomic analyses employing a bacterial GWAS approach. Notably, our findings reveal a strong influence of the capsular serotype compared to the genetic background, and the absence of specific individual genomic loci modulating pneumococcal growth kinetics, particularly the *in vitro* growth rate. These findings may partially explain why some serotypes and lineages colonise and cause invasive diseases better than others and may have implications for developing measures to prevent and control pneumococcal infections.

Our findings suggest that the capsular serotype and genetic background influence pneumococcal growth kinetics. The findings are generally consistent with results from experimental studies by others[13–15], which used capsule-switched strains to exclude the impact of the genetic background and showed an alternation of growth dynamics due to the capsule, although there may be potential differences in the growth media and conditions. Specifically, encapsulated strains showed higher growth rates than unencapsulated strains in nutrient-rich conditions[14]. These findings suggest that the capsule is crucial for pneumococcal growth, but its effect depends on nutrient availability – promoting and interfering with growth when nutrients are abundant and limited, respectively[14,27]. A recent study by Tóthpál et al.[15] suggested that the serotype partially explains pneumococcal growth and suggested the genetic backgrounds, which vary in their ability to colonise the nasopharynx and cause invasive diseases[4,46–48], may also play a role. However, the authors did not fully quantify the impact of the genetic background since the isolates were not sequenced[15]. Our study confirms the strong influence of both the pneumococcal capsular serotype and genetic background or lineage on the growth kinetics of the isolates. The influence of the serotype appeared to be higher than the genetic background for the average growth rate, while the other growth features appeared to show an opposite trend. These findings

299 suggest that the respective influence of the capsular serotype and lineage varies by specific
300 derived pneumococcal growth features.
301
302 Interestingly, although our findings revealed that the capsular serotype and genetic background
303 influence pneumococcal growth kinetics, statistical analysis showed that the variance attributed
304 to the capsular serotype was higher than the genetic background. Specifically, the capsular
305 serotype exhibits a more pronounced growth rate, maximum growth change, and maximum
306 growth density than the genetic background. In contrast, the genetic background had a higher
307 effect on the lag phase duration than the serotype. We acknowledge that our comparison of
308 growth features in the same genetic background could also mean a large amount of genetic
309 variation for some of the lineages, some of which could be responsible for changes in growth.
310 These findings illustrate the importance of the capsular serotype and genetic background on
311 pneumococcal growth kinetics, which may partially explain their variable prevalence in carriage
312 and disease.
313
314 Previous studies have revealed individual bacterial genomic loci that impact other phenotypes,
315 including pathogenicity[49–57], virulence[58–61], host, niche, and environmental adaptation[52,54,62],
316 conjugation[63], and antimicrobial resistance[64–71], independent of the genetic background. Our
317 GWAS findings revealed no elevated signals in specific genomic loci, suggesting that individual
318 variants may not significantly influence pneumococcal growth kinetics independent of the
319 genetic background and serotype. Previous studies have shown that mutations in other capsule
320 biosynthesis genes (*wchA* or *cpsE*) encoding a protein that links activated glucose phosphates to
321 lipid carriers[44] also influence the growth dynamics of the pneumococcus[43]. However, our results
322 are contrary to our speculation that genetic variation in the capsule polysaccharide synthesis
323 genes would be strongly associated with pneumococcal growth features. These findings suggest
324 that specific genomic loci in the pneumococcal genome do not strongly influence pneumococcal
325 growth kinetics, hinting at the potential impact of the overall genetic background and several loci
326 possibly with small effects individually. While it's possible that such follow-up studies are

327 needed to provide additional support for the importance of these variants on pneumococcal
328 growth kinetics.
329
330 In contrast to findings in a previous study[14], we found no robust evidence for the association
331 between the *in vitro* growth rates and the population-level prevalence of the serotypes in
332 nasopharyngeal carriage. However, this may have occurred as a result of a limitation of our
333 study, since we studied a collection of randomly sampled invasive isolates, and the growth
334 experiments were conducted in conditions similar to that of blood, while the Hathaway et al.
335 study[14] used a non-random selection of both carriage and disease isolates. It is likely that isolates
336 have adaptations to the environmental conditions in which they were isolated, and so *in vitro*
337 growth characteristics for any given isolate may be different under experimental conditions
338 simulating different niches. In addition, the *in vitro* growth environment cannot fully represent *in*
339 *vivo* conditions and fails to consider the impact of selective pressures coming from the serotype-
340 specific responses (antibody) or non-specific responses (phagocytic) and potential interactions
341 with other pathogens. Presumably, the lack of association between the invasiveness and
342 prevalence of serotypes in invasive disease and carriage, and the growth rates could reflect that
343 the conditions in the *in vitro* growth environment do not fully represent *in vivo* conditions.
344 Collectively, these findings may suggest that these and other factors may substantially influence
345 the complex pneumococcal disease pathogenesis *in vivo*.
346
347 The strength of this study is the analysis of an extensive collection of well-phenotyped natural
348 isolates and the combination of phenotyping, and estimation of growth features using statistical
349 models and statistical genetics to determine the association between the growth features and
350 genetic variation across the entire genome. As previously noted by Tothpal et al.[15], the culturing
351 history of the isolates may impact the growth kinetics. In our study, we consistently cultured all
352 the isolates after isolation from patients using standardised culturing procedures (single assay,
353 identical initial concentration, environmental conditions or temperature, and equipment to
354 measure the optical density and repeat measurements to spot technical variability[13]). Compared
355 to the present study, future studies would benefit from larger sample sizes and may find more
356 results as we only explained some of the heritability of the growth features. We also
357 acknowledge that the methods used to estimate heritability may also sometimes perform poorly

358 on some bacteria species, therefore, more robust methods are needed to further validate our
359 findings[72]. Because we conducted our experiments for a relatively short duration (~15 hours), we
360 could not robustly characterise the kinetics at the stationary phase due to incomplete data or the
361 observed high variability of its kinetics. Therefore, we recommend increasing the culturing time
362 from 15 to >24 hours in follow-up studies, as done elsewhere[14,15], to characterise the stationary
363 phase kinetics of the cultured isolates robustly. Furthermore, since the pneumococcus invades
364 different niches and systemic tissues with varying conditions, including oxygen concentration
365 and temperature[12,14–16], future studies should investigate the effect of these environmental
366 conditions on pneumococcal growth and identify pneumococcal genomic variants associated
367 with growth kinetics for each condition. Considering that our dataset contained only a third of
368 the >100 known pneumococcal serotypes globally[3]; due to the geographical variation in the
369 distribution of pneumococcal serotypes[73], future studies should investigate all serotypes to
370 provide further insights on the impact of different serotypes in pneumococcal growth kinetics.
371 Additionally, considering that we generated the phylogeny of the entire pneumococcal species
372 rather than isolates belonging to a single strain or lineage, as defined by several
373 approaches[34,74,75], we did not remove recombinations using widely used methods designed for
374 clonal bacterial populations, such as Gubbins[76]. Therefore, there is a possibility that some branch
375 lengths in the phylogeny used for the ancestral trait reconstruction analysis may be less accurate
376 although the topology is usually robust to recombination[77]. Furthermore, there may be other
377 variants that potentially influence the pneumococcal growth kinetics which are regarded as not
378 statistically significant in this study based on our conservative genome-wide significance
379 threshold. As more and more similar studies are conducted, particularly those utilising much
380 larger datasets, potential variants associated with pneumococcal growth kinetics, may be
381 discovered.
382
383 Our analysis of population-level genomic and high-throughput *in vitro* pneumococcal growth
384 data suggests that the capsular serotype and lineage influence pneumococcal growth. However,
385 although some growth features are highly heritable, there appears to be no specific genetic
386 variants which strongly influence these phenotypes suggesting that a combination of genomic
387 loci with small effects individually or the entire genetic background may play a greater role in
388 pneumococcal growth kinetics. Overall, our study provides a proof-of-concept for combining

389 phenotyping and population genomics to understand the genetic basis for pneumococcal growth,
390 an approach which can be applied to study other bacteria of clinical significance and those used
391 in bioengineering, genetics, and the food industry[78–80]. Ultimately, we anticipate that applying
392 our approach will have implications for devising strategies to reduce life-threatening bacterial
393 infections and improve food production and safety.

394 Methods

395 Ethics statement

396 The isolates used in this work were collected through an observational cohort study
397 Pneumococcal Bacteraemia Collection Nijmegen (PBCN)[33]. The non-applicability of the
398 Medical Research Involving Human Subjects Act (WMO) on the study design was confirmed by
399 the Regional Ethics Committee, and the local medical ethics committees of the participating
400 hospitals in the Netherlands approved the study procedures.

401 Sample characteristics and whole-genome sequencing

402 Three hundred and forty-eight (n=348) pneumococcal isolates collected from patients with
403 invasive pneumococcal diseases through the PBCN observational cohort study[33] were available
404 for the study. We included a single isolate per patient as invasive pneumococcal infections are
405 typically clonal, so a single isolate is sufficient. The consecutive blood isolates were collected at
406 two hospitals in Nijmegen, the Netherlands, before (2000-2006) and after (2007-2011) the
407 introduction of the PCV7 vaccine into the Dutch paediatric immunisation programme[33]. Most of
408 the bacteremia cases (97%) concerned adults. Additional information on the PBCN study and
409 collected clinical variables have been previously described[33,81]. To generate whole genome
410 sequences, we sequenced a single isolate per patient using an Illumina HiSeq sequencer
411 (Illumina, San Diego, CA, USA). We then assembled the generated reads using SPAdes genome
412 assembler (version 3.14.0)[82], and evaluated the draft assemblies using assembly-stats (version:
413 1.0.1) (https://github.com/sanger-pathogens/assembly-stats). In addition, we used a combination

414 of metrics, including genome size and the number of contigs, to assess the quality of the draft
415 assemblies for inclusion in the analysis.

416 Measurement of *in vitro* pneumococcal growth kinetics and derivation of 417 growth features

418 We measured the pneumococcal growth kinetics for slightly over 15 hours as described by
419 Arends et al.[13]. In short, we incubated a frozen aliquot of standardised pneumococcal inoculum
420 in 1.5 ml of blood-like medium (50% M17, 50% casamino acids tryptone [CAT] medium, 0.25%
421 glucose). We measured growth kinetics at 37°C and 5% CO_2 every 10 min for over 15 hours at
422 OD_{620} with use of a humidity cassette by microplate reader (Spark 10M, Tecan, Switzerland).
423 We added 15 µL of inoculum was added to 1.5 mL of prewarmed rich growth medium (45%
424 M17, 45% CAT, 0.225% glucose, 10% fetal calf serum [Greiner Bio-one], 26 U/mL catalase
425 [Sigma-Aldrich C1345]) in each well of a sterile flat-bottomed 48-well plate (Nunclon Surface,
426 Nunc, Denmark). The growth medium was supplemented with catalase. We chose these nutrient-
427 rich conditions to bring forward differences in the growth features of the pneumococcus. We
428 performed six repeat measurements on three separate days for each isolate. To understand the
429 growth kinetics of the invasive Dutch pneumococcal isolates, we fitted a modified logistic
430 growth curve function, i.e., $L_0 + L_{max}/(1 + e^{-r(t-t_0)})$, separately to the growth data for each
431 isolate for each replicate, using the "optim" function in stats (version 4.3.2) R package
432 (https://www.R-project.org/). We then averaged the inferred values for each replicate to obtain
433 average estimates for each parameter. The parameter L_0 captures the initial cell density measured
434 as absorbance or optical density at a wavelength of 620 nm (OD_{620}); L_{max} is the maximum cell
435 density; t_0 is the initial time, and r is the average growth rate. We fitted two curves in total for
436 each isolate based on the averaged data from the six replicates using the limited-memory
437 Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B) optimisation algorithm by minimising the
438 residual sum of squares[83]. We fitted the first curve on all the data points and the second curve on
439 the data measured until the start of the end of the stationary phase or the beginning of the death
440 phase, whichever came earlier. We defined the end of the stationary phase or the start of the
441 death phase as the time when the optical cell density decreased by >5% from the maximum cell
442 density. The curve fitted based on the data measurements collected up to the end of the stationary
443 phase or death phase, or the end of the experiment, if none of these phases had been reached,

444 captured the growth dynamics better than the curve fitted using all the data points to accurately
445 estimate the growth rate and the maximum cell density corresponding to the peak of the
446 pneumococcal growth curve, especially for growth curves with observably short stationary
447 phases and rapid autolysis. Therefore, parameters for the downstream analyses were inferred
448 from the curves fitted to the reduced dataset.
449
450 We determined four growth phenotypes for each pneumococcal isolate estimated from the fitted
451 growth curve, namely the lag-phase duration, maximum growth density (H_{max}), i.e., captured by
452 parameter L_{max}, maximum growth change (ΔH) corresponding to the difference of the maximum
453 and minimum density (L_{max}-L_0), and average growth rate (r). We defined the lag-phase duration
454 as the time when the optical cell density was at least 1.5-fold higher than the starting initial cell
455 density, as described by Arends and colleagues[13]. Due to the challenges of characterising the
456 kinetics at the stationary or death phase, partly due to the high strain-to-strain variability or short
457 study period, we did not derive the growth features capturing these phases, including the duration
458 of the stationary and death phases. The logistic growth curves showed an excellent fit for the
459 data, which provided an opportunity to investigate the impact of capsular serotype and genetic
460 background or lineages and specific genomic loci, independent of the genetic background, on the
461 estimated growth parameters or features.

462 Molecular typing of sequenced isolates

463 Molecular typing of the isolates to determine capsular serotypes and sequence types (STs) was
464 done *in silico* using Pathogenwatch (https://pathogen.watch/), a global platform for genomic
465 surveillance[84]. We determined the capsular serotypes and STs using SeroBA (version v1.0.1)[85]
466 and the multilocus sequence typing (MLST) scheme for the pneumococcus[35], respectively.
467 Pneumococcal lineages were defined based on the global pneumococcal sequence cluster
468 (GPSC) nomenclature[34] using PopPUNK (version 1.1.0)[86]. In addition, we checked the species of
469 each sequenced isolate using Speciator (version 3.0.1) in Pathogenwatch.

470 Construction of multiple sequence alignments and phylogenetic tree

471 The generated draft assemblies for each isolate were mapped to the pneumococcal ATCC
472 700669 reference genome (GenBank accession: NC_011900) using Snippy (version 4.6.0)

473 (https://github.com/tseemann/snippy). We identified 155,305 SNPs in the generated alignment
474 using SNP-sites (version 2.3.2)[87] and then used them to construct a maximum-likelihood
475 phylogeny using IQ-TREE (version 2.0.3)[88]. We used the alignment of variable sites by mapping
476 to a reference genome to generate a maximum-likelihood phylogeny using IQ-TREE as it has
477 been to produce more accurate phylogenies[89]. The general time reversible (GTR) nucleotide
478 substitution model and Gamma heterogeneity of the substitution rates among sites. We assessed
479 branch support with ultrafast bootstrap approximation using "-B 1000" option[90]. We optimised
480 bootstrapped trees by nearest neighbour interchange on bootstrap alignment by specifying the "--
481 bnni" option. We rooted the phylogenetic tree of the pneumococcal isolates using *Streptococcus*
482 *oralis* as an outgroup (not shown in the phylogeny). An annotated and interactive version of the
483 phylogeny is available on the Microreact web tool[91]
484 (https://microreact.org/project/wxmdblfgprepwugvbbtaq7).

485 Quantifying the phylogenetic signal

486 To assess the correlation between the estimated growth features and pneumococcal phylogeny,
487 we used models of continuous character evolution to estimate the phylogenetic signal using
488 Pagel's λ[38] implemented in the "phylosig" function in phytools package (version 2.1.1)[92]. Pagel's
489 λ transforms the internal branches relative to the terminal branches and generates values ranging
490 from zero to one, with smaller values indicating no phylogenetic signal or independent evolution
491 of the phenotype from the phylogeny. In comparison, higher values close to one show a high
492 phylogenetic signal or a decreased phenotype variability among genetically similar taxa at the
493 tips of the phylogeny. First, we transformed the estimated growth features to a normal
494 distribution using a rank-based inverse normal transform implemented in the RNOmni package
495 (version 1.0.1.2) (https://cran.r-project.org/web/packages/RNOmni/). Next, we mapped the
496 estimated continuous growth feature values onto the phylogeny using fast estimation of
497 maximum likelihood for ancestral states[92].

498 Bacterial GWAS

499 We identified the presence and absence patterns of unitig sequences To capture genetic variation
500 in the pneumococcal genomes included in this study for the GWAS analysis. We identified
501 unitig sequences from 31bp *k*-mers using Bifrost (version 1.0.6)[93]. First, we inferred unitig

502 sequences from a compacted De Bruijn graph constructed using *k*-mers of length 31 using draft
503 assemblies of all the pneumococcal isolates included in this study. We then generated De Bruijn
504 graphs for each isolate and queried them with the unitig sequences found in the entire dataset to
505 identify the presence and absence of each unitig found in each sequenced genome. We
506 considered a unitig present when 100% of the *k*-mer was presented in the De Bruijn graph. We
507 combined the presence and absence patterns of the unitig sequences from all the isolates to
508 construct a single matrix.
509
510 Since the linear mixed effects models require the tested continuous phenotype to be normally
511 distributed, we used the normalised growth features after applying the rank-based inverse normal
512 transformation using the RNOmni package (version 1.0.1.2) (https://cran.r-
513 project.org/web/packages/RNOmni/). We performed GWAS analysis of the pneumococcal
514 growth features using robust linear mixed effects models implemented in Pyseer (version 1.3.7-
515 dev)[94]. A kinship matrix specifying the variances and covariances of the random effects for each
516 isolate was generated using the "similarity_pyseer" script in Pyseer. Using this kinship matrix,
517 Pyseer fits an unobserved random effect for each isolate to account for the clonal population
518 structure in the GWAS. We excluded unitigs with <5% minimum allele frequency and >95%
519 maximum allele frequency. The inferred statistical significance for each variant was adjusted to
520 correct for multiple testing using the Bonferroni correction. We estimated the total number of
521 unique unitig presence and absence patterns in the dataset using a script included in Pyseer
522 (https://github.com/mgalardini/pyseer/blob/master/scripts/count_patterns.py), and used the
523 obtained value when adjusting for multiple testing using the Bonferroni correction approach.
524 Therefore, we defined the genome-wide statistical significance threshold as $P < 2.58\times10^{-07}$, i.e.,
525 α/G, where α=0.05 and G=194,053 is the number of unique unitig presence and absence patterns
526 while the suggestive threshold as $P<5.15\times10^{-06}$, whereby α=1. To check potential issues arising
527 from the population structure, we generated quantile-quantile (QQ) plots from the output of
528 GWAS analysis for each growth feature to compare the expected and observed *P*-values using
529 qqman (version 0.1.7)[95]. The overall proportion of phenotypic variability explained by variation
530 in the genome (narrow-sense heritability [h^2]) was estimated using GEMMA (version 0.98.1)[96]
531 and Pyseer. The input files for GEMMA were prepared based on the unitig presence and absence

532 data using PLINK (version 1.90b4) [97]. To identify variants associated with the growth features
533 independently of the serotype, we repeated the GWAS by including the serotype as a covariate.

534 Annotation of variants identified by the GWAS analysis

535 To annotate the unitig sequences with statistical significance values below the suggestive and
536 genome-wide significance threshold, we used a custom BioPython [98] wrapper script for
537 BLASTN (version 2.5.0+)[99] to identify genomic features in the ATCC 700669 reference genome
538 (GenBank accession: NC_011900) and other complete pneumococcal genomes downloaded
539 from GenBank[99]. The GWAS results were summarised and visualised using Manhattan plots
540 using R (version 4.0.3) [https://www.R-project.org/].

541 Other statistical analysis

542 We compared the growth features between serotypes and lineages using the Kruskal-Wallis non-
543 parametric test. The distribution of the growth features was plotted using violin plots using the
544 vioplot package [version 0.4.0] (https://CRAN.R-project.org/package=vioplot). To assess the
545 contribution of serotype and lineage to the variability in the pneumococcal growth features, we
546 used linear mixed-effects models implemented in the lme4 package (version 1.1.35.1)[100] were
547 used, with each growth feature specified as the outcome and the serotype and lineage as random
548 effects. The statistical significance for the linear mixed model was inferred using the lmerTest
549 package [version 3.1.3] (https://CRAN.R-project.org/package=lmerTest). To assess the
550 contribution of the serotype and lineage to the variability of each growth feature, we calculated
551 the proportion of variance for the serotype and lineage components to the total variance. We
552 calculated the average invasiveness of the serotypes based on a meta-analysis of data collected in
553 different countries in infants before the introduction of the PCVs[47] using the random-effects
554 model implemented in the metafor package [version 2.4.0] (https://CRAN.R-
555 project.org/package=metafor). We defined invasiveness as the odds ratio of detecting a
556 pneumococcal serotype among isolates causing invasive disease compared to nasopharyngeal
557 carriage. We added a value of one to cells with zero counts. Due to the unavailability of the
558 nasopharyngeal carriage data, we imputed the prevalence of the serotypes in the carriage by
559 dividing the prevalence in invasive disease by the average invasiveness of each serotype. Next,
560 we used a Spearman correlation test to assess the association between the mean prevalence of

561 serotypes in the nasopharyngeal carriage and invasive disease and average invasiveness with the
562 average *in vitro* pneumococcal growth rate.

563 Acknowledgements

564 The authors would like to thank the study participants, guardians, and the clinical and laboratory
565 staff who collected and processed the samples at various laboratories in the Netherlands. We
566 would also like to acknowledge the support of the sequencing, core, parasites and microbes
567 teams, the Bentley lab at the Wellcome Sanger Institute, and the Weinberger and Pitzer labs at
568 the Yale School of Public Health for their feedback on the analysis. We acknowledge the
569 participating hospitals and affiliated researchers for their contributions and support.

570 Funding information

571 This study was supported by the TARGET project (grant number: JPIAMR2019-087) funded by
572 JPI-AMR–ZonMW, The Netherlands, and by the SNSF Swiss Postdoctoral Fellowship (SNSF
573 grant number: TMPFP3_209768). Activities at the Wellcome Sanger Institute were funded by
574 the Bill and Melinda Gates Foundation (grant number: OPP1034556) and Wellcome Trust
575 (2016-2021 core award grant number: 206194). The funders had no role in study design, data
576 collection, analysis, the decision to publish, and manuscript preparation. The findings do not
577 necessarily reflect the official views and policies of the author's institutions and funders. For the
578 purpose of open access, the authors have applied a CC BY public copyright licence to any
579 Author Accepted Manuscript version arising from this submission.

580 Author contributions

581 C.C., M.I.dJ., and A.J.H.C. conceived and designed the study. M.Id.J., D.W.A., I.H.P., and
582 A.J.H.C. collected and processed the pneumococcal isolates. S.W.L. and S.D.B. performed
583 whole-genome sequencing, quality control and genomic analysis. C.C. performed statistical and
584 bioinformatic analyses. D.W.A. contributed to data analysis and interpretation. M.I.dJ. and
585 A.J.H.C. contributed to data interpretation. C.C., D.W.A., M.I.dJ., and A.J.H.C. wrote the first

586 draft of the manuscript. C.C., D.W.A., S.W.L., I.H.P., A.Y., J.A.L., A.L.W., D.M.W., S.D.B.,
587 M.I.dJ., and A.J.H.C. reviewed and approved the manuscript.

588 Declaration of interests

589 The authors declare no competing interests.

590 References

591 1. GBD 2019 Antimicrobial Resistance Collaborators. Global mortality associated with 33
592 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study
593 2019. *Lancet* (2022) doi:10.1016/S0140-6736(22)02185-7.
594 2. von Gottberg, A. *et al.* Effects of vaccination on invasive pneumococcal disease in South
595 Africa. *N. Engl. J. Med.* **371**, 1889–1899 (2014).
596 3. Ganaie, F. *et al.* A New Pneumococcal Capsule Type, 10D, is the 100th Serotype and Has a
597 Large *cps* Fragment from an Oral Streptococcus. *mBio* vol. 11 Preprint at
598 https://doi.org/10.1128/mbio.00937-20 (2020).
599 4. Brueggemann, A. B. *et al.* Clonal relationships between invasive and carriage Streptococcus
600 pneumoniae and serotype- and clone-specific differences in invasive disease potential. *J.*
601 *Infect. Dis.* **187**, 1424–1432 (2003).
602 5. Grabenstein, J. D. & Musey, L. K. Differences in serious clinical outcomes of infection
603 caused by specific pneumococcal serotypes among adults. *Vaccine* **32**, 2399–2405 (2014).
604 6. Müller, A. *et al.* Streptococcus pneumoniae Serotypes Associated with Death, South Africa,
605 2012-2018. *Emerg. Infect. Dis.* **28**, 166–179 (2022).
606 7. Cohen, C. *et al.* Streptococcus pneumoniae Serotypes and Mortality in Adults and
607 Adolescents in South Africa: Analysis of National Surveillance Data, 2003 - 2008. *PLoS*
608 *One* **10**, e0140185 (2015).

609 8. Donati, C. *et al.* Structure and dynamics of the pan-genome of Streptococcus pneumoniae
610 and closely related species. *Genome Biol.* **11**, R107 (2010).
611 9. Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. The role of Streptococcus
612 pneumoniae virulence factors in host respiratory colonization and disease. *Nat. Rev.*
613 *Microbiol.* **6**, 288–301 (2008).
614 10. Hathaway, L. J., Grandgirard, D., Valente, L. G., Täuber, M. G. & Leib, S. L. Streptococcus
615 pneumoniae capsule determines disease severity in experimental pneumococcal meningitis.
616 *Open Biol.* **6**, (2016).
617 11. Hamaguchi, S., Zafar, M. A., Cammer, M. & Weiser, J. N. Capsule Prolongs Survival of
618 Streptococcus pneumoniae during Starvation. *Infect. Immun.* **86**, (2018).
619 12. Im, H. *et al.* Anatomical Site-Specific Carbohydrate Availability Impacts Streptococcus
620 pneumoniae Virulence and Fitness during Colonization and Disease. *Infect. Immun.* **90**,
621 e0045121 (2022).
622 13. Arends, D. W. *et al.* Differential Pneumococcal Growth Features in Severe Invasive Disease
623 Manifestations. *Microbiol Spectr* **10**, e0005022 (2022).
624 14. Hathaway, L. J. *et al.* Capsule type of Streptococcus pneumoniae determines growth
625 phenotype. *PLoS Pathog.* **8**, e1002574 (2012).
626 15. Tóthpál, A., Desobry, K., Joshi, S. S., Wyllie, A. L. & Weinberger, D. M. Variation of
627 growth characteristics of pneumococcus with environmental conditions. *BMC Microbiol.*
628 **19**, 304 (2019).
629 16. Small, P. M., Täuber, M. G., Hackbarth, C. J. & Sande, M. A. Influence of body
630 temperature on bacterial growth rates in experimental pneumococcal meningitis in rabbits.
631 *Infect. Immun.* **52**, 484–487 (1986).

632 17. Klumpp, S., Zhang, Z. & Hwa, T. Growth Rate-Dependent Global Effects on Gene
633 Expression in Bacteria. *Cell* **139**, 1366–1375 (2009).
634 18. Nilsson, G., Belasco, J. G., Cohen, S. N. & von Gabain, A. Growth-rate dependent
635 regulation of mRNA stability in Escherichia coli. *Nature* **312**, 75–77 (1984).
636 19. Rogers, A. H., Pilowsky, K. & Zilm, P. S. The effect of growth rate on the adhesion of the
637 oral bacteria Streptococcus mutans and Streptococcus milleri. *Arch. Oral Biol.* **29**, 147–150
638 (1984).
639 20. Russell, J. B., Sharp, W. M. & Baldwin, R. L. The effect of pH on maximum bacterial
640 growth rate and its possible role as a determinant of bacterial competition in the rumen. *J.*
641 *Anim. Sci.* **48**, 251–255 (1979).
642 21. Lin-Chao, S. & Bremer, H. Effect of the bacterial growth rate on replication control of
643 plasmid pBR322 in Escherichia coli. *Mol. Gen. Genet.* **203**, 143–149 (1986).
644 22. Nabergoj, D., Modic, P. & Podgornik, A. Effect of bacterial growth rate on bacteriophage
645 population growth rate. *Microbiologyopen* **7**, e00558 (2018).
646 23. Łapińska, U. *et al.* Fast bacterial growth reduces antibiotic accumulation and efficacy. *Elife*
647 **11**, (2022).
648 24. Smirnova, G. V. & Oktyabrsky, O. N. Relationship between Escherichia coli growth rate
649 and bacterial susceptibility to ciprofloxacin. *FEMS Microbiol. Lett.* **365**, (2018).
650 25. Tuomanen, E., Cozens, R., Tosch, W., Zak, O. & Tomasz, A. The rate of killing of
651 Escherichia coli by beta-lactam antibiotics is strictly proportional to the rate of bacterial
652 growth. *J. Gen. Microbiol.* **132**, 1297–1304 (1986).
653 26. Grant, A. J. *et al.* Bacterial growth rate and host factors as determinants of intracellular
654 bacterial distributions in systemic Salmonella enterica infections. *Infect. Immun.* **77**, 5608–

655 5611 (2009).
656 27. Müller, A. *et al.* Pneumococcal serotype determines growth and capsule size in human
657 cerebrospinal fluid. *BMC Microbiol.* **20**, 16 (2020).
658 28. Gulig, P. A. & Doyle, T. J. The Salmonella typhimurium virulence plasmid increases the
659 growth rate of salmonellae in mice. *Infect. Immun.* **61**, 504–511 (1993).
660 29. Shlla, B. *et al.* The Rgg1518 transcriptional regulator is a necessary facet of sugar
661 metabolism and virulence in Streptococcus pneumoniae. *Mol. Microbiol.* **116**, 996–1008
662 (2021).
663 30. Zhang, C. *et al.* MetR is a molecular adaptor for pneumococcal carriage in the healthy upper
664 airway. *Mol. Microbiol.* **116**, 438–458 (2021).
665 31. Gao, Y. & Li, H. Quantifying and comparing bacterial growth dynamics in multiple
666 metagenomic samples. *Nat. Methods* **15**, 1041–1044 (2018).
667 32. Korem, T. *et al.* Growth dynamics of gut microbiota in health and disease inferred from
668 single metagenomic samples. *Science* **349**, (2015).
669 33. Cremers, A. J. H. *et al.* Effects of 7-valent pneumococcal conjugate 1 vaccine on the
670 severity of adult 2 bacteremic pneumococcal pneumonia. *Vaccine* **32**, 3989–3994 (2014).
671 34. Gladstone, R. A. *et al.* International genomic definition of pneumococcal lineages, to
672 contextualise disease, antibiotic resistance and vaccine impact. *EBioMedicine* **43**, 338–346
673 (2019).
674 35. Enright, M. C. & Spratt, B. G. A multilocus sequence typing scheme for Streptococcus
675 pneumoniae: identification of clones associated with serious invasive disease. *Microbiology*
676 **144 (Pt 11)**, 3049–3060 (1998).
677 36. Croucher, N. J. *et al.* Rapid pneumococcal evolution in response to clinical interventions.

678 *Science* **331**, 430–434 (2011).
679 37. Croucher, N. J. *et al.* Selective and genetic constraints on pneumococcal serotype switching.
680 *PLoS Genet.* **11**, e1005095 (2015).
681 38. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* vol. 401 877–884
682 Preprint at https://doi.org/10.1038/44766 (1999).
683 39. Jaillard, M. *et al.* A fast and agnostic method for bacterial genome-wide association studies:
684 Bridging the gap between k-mers and genetic events. *PLoS Genet.* **14**, e1007758 (2018).
685 40. Lees, J. A. *et al.* Improved Prediction of Bacterial Genotype-Phenotype Associations Using
686 Interpretable Pangenome-Spanning Regressions. *MBio* **11**, (2020).
687 41. Power, R. A., Parkhill, J. & de Oliveira, T. Microbial genome-wide association studies:
688 lessons from human GWAS. *Nat. Rev. Genet.* **18**, 41–50 (2017).
689 42. Rodenburg, G. D. *et al.* Effects of pneumococcal conjugate vaccine 2 years after its
690 introduction, the Netherlands. *Emerg. Infect. Dis.* **16**, 816–823 (2010).
691 43. Schaffner, T. O. *et al.* A point mutation in cpsE renders Streptococcus pneumoniae
692 nonencapsulated and enhances its growth, adherence and competence. *BMC Microbiol.* **14**,
693 210 (2014).
694 44. Bentley, S. D. *et al.* Genetic analysis of the capsular biosynthetic locus from all 90
695 pneumococcal serotypes. *PLoS Genet.* **2**, e31 (2006).
696 45. Croucher, N. J. *et al.* Population genomics of post-vaccine changes in pneumococcal
697 epidemiology. *Nat. Genet.* **45**, 656–663 (2013).
698 46. Hanage, W. P. *et al.* Invasiveness of serotypes and clones of Streptococcus pneumoniae
699 among children in Finland. *Infect. Immun.* **73**, 431–435 (2005).
700 47. Colijn, C., Corander, J. & Croucher, N. J. Designing ecologically optimized pneumococcal

701 vaccines using population genomics. *Nat Microbiol* **5**, 473–485 (2020).

702 48. Massora, S. *et al.* Invasive disease potential of Streptococcus pneumoniae serotypes before
703 and after 10-valent pneumococcal conjugate vaccine introduction in a rural area, southern
704 Mozambique. *Vaccine* **37**, 7470–7477 (2019).

705 49. Lees, J. A. *et al.* Joint sequencing of human and pathogen genomes reveals the genetics of
706 pneumococcal meningitis. *Nat. Commun.* **10**, 2176 (2019).

707 50. Li, Y. *et al.* Genome-wide association analyses of invasive pneumococcal isolates identify a
708 missense bacterial mutation associated with meningitis. *Nat. Commun.* **10**, 178 (2019).

709 51. Young, B. C. *et al.* Panton–Valentine leucocidin is the key determinant of Staphylococcus
710 aureus pyomyositis in a bacterial GWAS. *eLife* vol. 8 Preprint at
711 https://doi.org/10.7554/elife.42486 (2019).

712 52. Chaguza, C. *et al.* Prophage-encoded immune evasion factors are critical for host infection,
713 switching, and adaptation. *Cell Genom* **2**, (2022).

714 53. Davies, M. R. *et al.* Atlas of group A streptococcal vaccine candidates compiled using
715 large-scale comparative genomics. *Nat. Genet.* **51**, 1035–1043 (2019).

716 54. Chewapreecha, C. *et al.* Genetic variation associated with infection and the environment in
717 the accidental pathogen. *Commun Biol* **2**, 428 (2019).

718 55. Peters, S. *et al.* Campylobacter jejuni genotypes are associated with post-infection irritable
719 bowel syndrome in humans. *Commun Biol* **4**, 1015 (2021).

720 56. Biggel, M. *et al.* Horizontally acquired papGII-containing pathogenicity islands underlie the
721 emergence of invasive uropathogenic Escherichia coli lineages. *Nat. Commun.* **11**, 5968
722 (2020).

723 57. Wee, B. A. *et al.* Population analysis of Legionella pneumophila reveals a basis for

724 resistance to complement-mediated killing. *Nat. Commun.* **12**, 7165 (2021).
725 58. Laabei, M. *et al.* Predicting the virulence of MRSA from its genome sequence. *Genome*
726 *Res.* **24**, 839–849 (2014).
727 59. Chaguza, C. *et al.* Bacterial genome-wide association study of hyper-virulent pneumococcal
728 serotype 1 identifies genetic variation associated with neurotropism. *Commun Biol* **3**, 559
729 (2020).
730 60. Boeck, L. *et al.* Mycobacterium abscessus pathogenesis identified by phenogenomic
731 analyses. *Nat Microbiol* **7**, 1431–1441 (2022).
732 61. Galardini, M. *et al.* Major role of iron uptake systems in the intrinsic extra-intestinal
733 virulence of the genus Escherichia revealed by a genome-wide association study. *PLoS*
734 *Genet.* **16**, e1009065 (2020).
735 62. Ma, K. C. *et al.* Adaptation to the cervical environment is associated with increased
736 antibiotic susceptibility in Neisseria gonorrhoeae. *Nat. Commun.* **11**, 4126 (2020).
737 63. Clark, R. R., Lapierre, P., Lasek-Nesselquist, E., Gray, T. A. & Derbyshire, K. M. A
738 Polymorphic Gene within the Mycobacterium smegmatis Locus Determines Mycobacterial
739 Self-Identity and Conjugal Compatibility. *MBio* **13**, e0021322 (2022).
740 64. Chewapreecha, C. *et al.* Comprehensive identification of single nucleotide polymorphisms
741 associated with beta-lactam resistance within pneumococcal mosaic genes. *PLoS Genet.* **10**,
742 e1004547 (2014).
743 65. Coll, F. *et al.* Genome-wide analysis of multi- and extensively drug-resistant
744 Mycobacterium tuberculosis. *Nat. Genet.* **50**, 307–316 (2018).
745 66. Creasy-Marrazzo, A. *et al.* Genome-wide association studies reveal distinct genetic
746 correlates and increased heritability of antimicrobial resistance in Vibrio cholerae under

747 anaerobic conditions. *Microb. Genom.* **8**, (2022).

748 67. Farhat, M. R. *et al.* Genomic analysis identifies targets of convergent positive selection in
749 drug-resistant Mycobacterium tuberculosis. *Nat. Genet.* **45**, 1183–1189 (2013).

750 68. Hicks, N. D., Carey, A. F., Yang, J., Zhao, Y. & Fortune, S. M. Bacterial Genome-Wide
751 Association Identifies Novel Factors That Contribute to Ethionamide and Prothionamide
752 Susceptibility in Mycobacterium tuberculosis. *MBio* **10**, (2019).

753 69. Farhat, M. R. *et al.* GWAS for quantitative resistance phenotypes in Mycobacterium
754 tuberculosis reveals resistance genes and regulatory regions. *Nat. Commun.* **10**, 2128
755 (2019).

756 70. Alam, M. T. *et al.* Dissecting vancomycin-intermediate resistance in staphylococcus aureus
757 using genome-wide association. *Genome Biol. Evol.* **6**, 1174–1185 (2014).

758 71. Mortimer, T. D., Zhang, J. J., Ma, K. C. & Grad, Y. H. Loci for prediction of penicillin and
759 tetracycline susceptibility in Neisseria gonorrhoeae: a genome-wide association study.
760 *Lancet Microbe* **3**, e376–e381 (2022).

761 72. Mallawaarachchi, S. *et al.* Genome-wide association, prediction and heritability in bacteria
762 with application to. *NAR Genom Bioinform* **4**, lqac011 (2022).

763 73. Johnson, H. L. *et al.* Systematic evaluation of serotypes causing invasive pneumococcal
764 disease among children under five: the pneumococcal global serotype project. *PLoS Med.* **7**,
765 (2010).

766 74. McGee, L. *et al.* Nomenclature of major antimicrobial-resistant clones of Streptococcus
767 pneumoniae defined by the pneumococcal molecular epidemiology network. *J. Clin.*
768 *Microbiol.* **39**, 2565–2571 (2001).

769 75. Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M. & Corander, J. Hierarchical and

770 spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* **30**,
771 1224–1228 (2013).
772 76. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial
773 whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
774 77. Hedge, J. & Wilson, D. J. Bacterial phylogenetic reconstruction from whole genomes is
775 robust to recombination but demographic inference is not. *MBio* **5**, e02158 (2014).
776 78. Tonner, P. D. *et al.* A Bayesian non-parametric mixed-effects model of microbial growth
777 curves. *PLoS Comput. Biol.* **16**, e1008366 (2020).
778 79. Yang, E. *et al.* Influence of culture media, pH and temperature on growth and bacteriocin
779 production of bacteriocinogenic lactic acid bacteria. *AMB Express* **8**, 10 (2018).
780 80. Alkema, W., Boekhorst, J., Wels, M. & van Hijum, S. A. F. T. Microbial bioinformatics for
781 food safety and production. *Briefings in Bioinformatics* vol. 17 283–292 Preprint at
782 https://doi.org/10.1093/bib/bbv034 (2016).
783 81. Cremers, A. J. H. *et al.* The Contribution of Genetic Variation of Streptococcus pneumoniae
784 to the Clinical Manifestation of Invasive Pneumococcal Disease. *Clinical Infectious*
785 *Diseases* Preprint at https://doi.org/10.1093/cid/ciy417 (2018).
786 82. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to
787 single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
788 83. Byrd, R. H., Lu, P., Nocedal, J. & Zhu, C. A limited memory algorithm for bound
789 constrained optimization. *SIAM J. Sci. Comput.* **16**, 1190–1208 (1995).
790 84. Argimón, S. *et al.* A global resource for genomic predictions of antimicrobial resistance and
791 surveillance of Salmonella Typhi at pathogenwatch. *Nat. Commun.* **12**, 2879 (2021).
792 85. Epping, L. *et al.* SeroBA: rapid high-throughput serotyping of Streptococcus pneumoniae

793 from whole genome sequence data. *Microb Genom* **4**, (2018).

794 86. Lees, J. A. *et al.* Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome*
795 *Res.* **29**, 304–316 (2019).

796 87. Page, A. J. *et al.* : rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb*
797 *Genom* **2**, e000056 (2016).

798 88. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and
799 effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol.*
800 *Evol.* **32**, 268–274 (2015).

801 89. Lees, J. A. *et al.* Evaluation of phylogenetic reconstruction methods using bacterial whole
802 genomes: a simulation based study. *Wellcome Open Res* **3**, 33 (2018).

803 90. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2:
804 Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

805 91. Argimón, S. *et al.* Microreact: visualizing and sharing data for genomic epidemiology and
806 phylogeography. *Microb Genom* **2**, e000093 (2016).

807 92. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other
808 things). *Methods in Ecology and Evolution* vol. 3 217–223 Preprint at
809 https://doi.org/10.1111/j.2041-210x.2011.00169.x (2012).

810 93. Holley, G. & Melsted, P. Bifrost: highly parallel construction and indexing of colored and
811 compacted de Bruijn graphs. *Genome Biol.* **21**, 249 (2020).

812 94. Lees, J. A., Galardini, M., Bentley, S. D., Weiser, J. N. & Corander, J. pyseer: a
813 comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* **34**,
814 4310–4312 (2018).

815 95. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan

816 plots. Preprint at https://doi.org/10.1101/005165.

817 96. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association
818 studies. *Nat. Genet.* **44**, 821–824 (2012).

819 97. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
820 linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

821 98. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular
822 biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

823 99. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
824 search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

825 100. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models
826 Usinglme4. *J. Stat. Softw.* **67**, (2015).

827 101. Croucher, N. J. *et al.* Role of conjugative elements in the evolution of the multidrug-
828 resistant pandemic clone Streptococcus pneumoniaeSpain23F ST81. *J. Bacteriol.* **191**,
829 1480–1489 (2009).

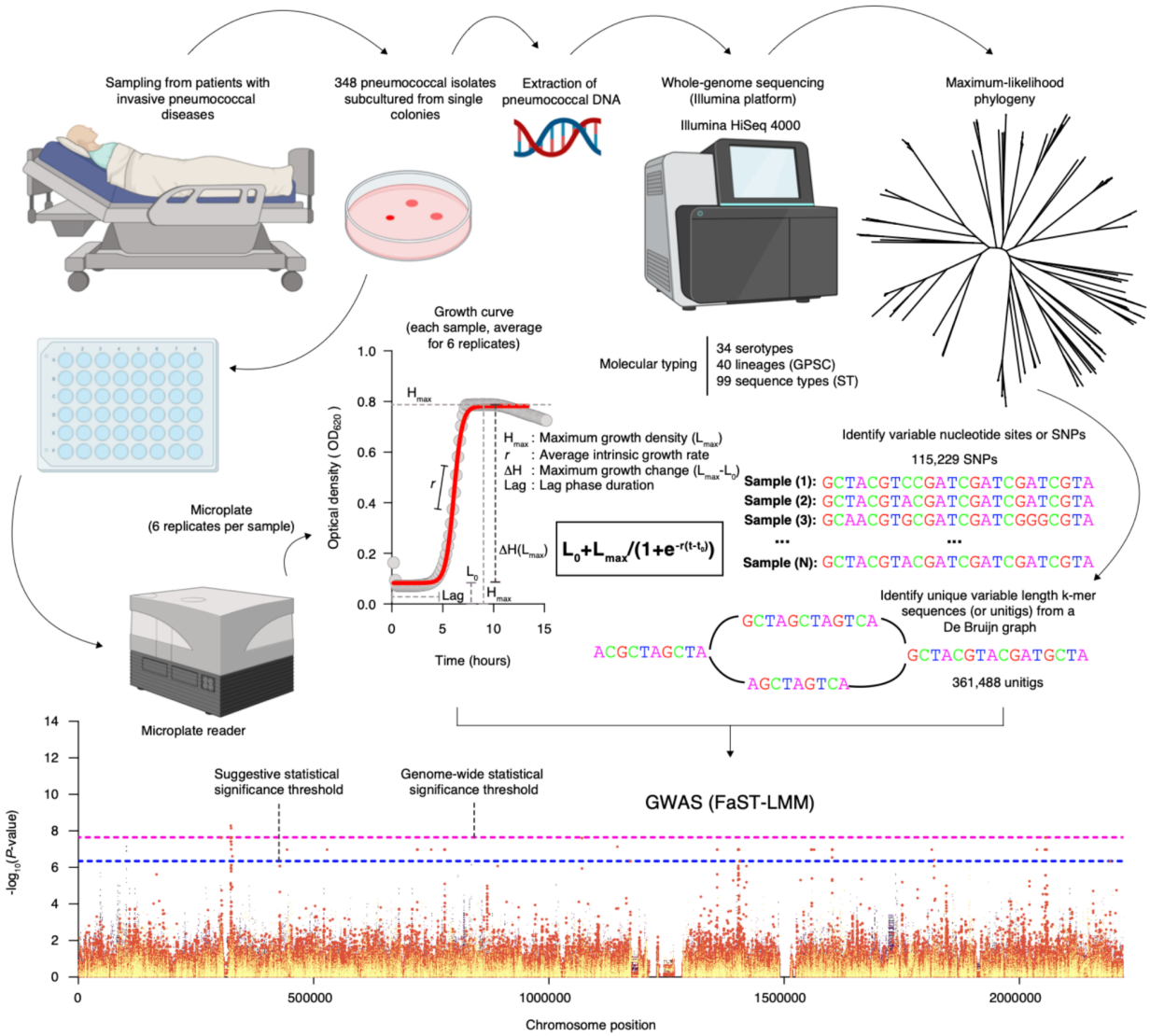
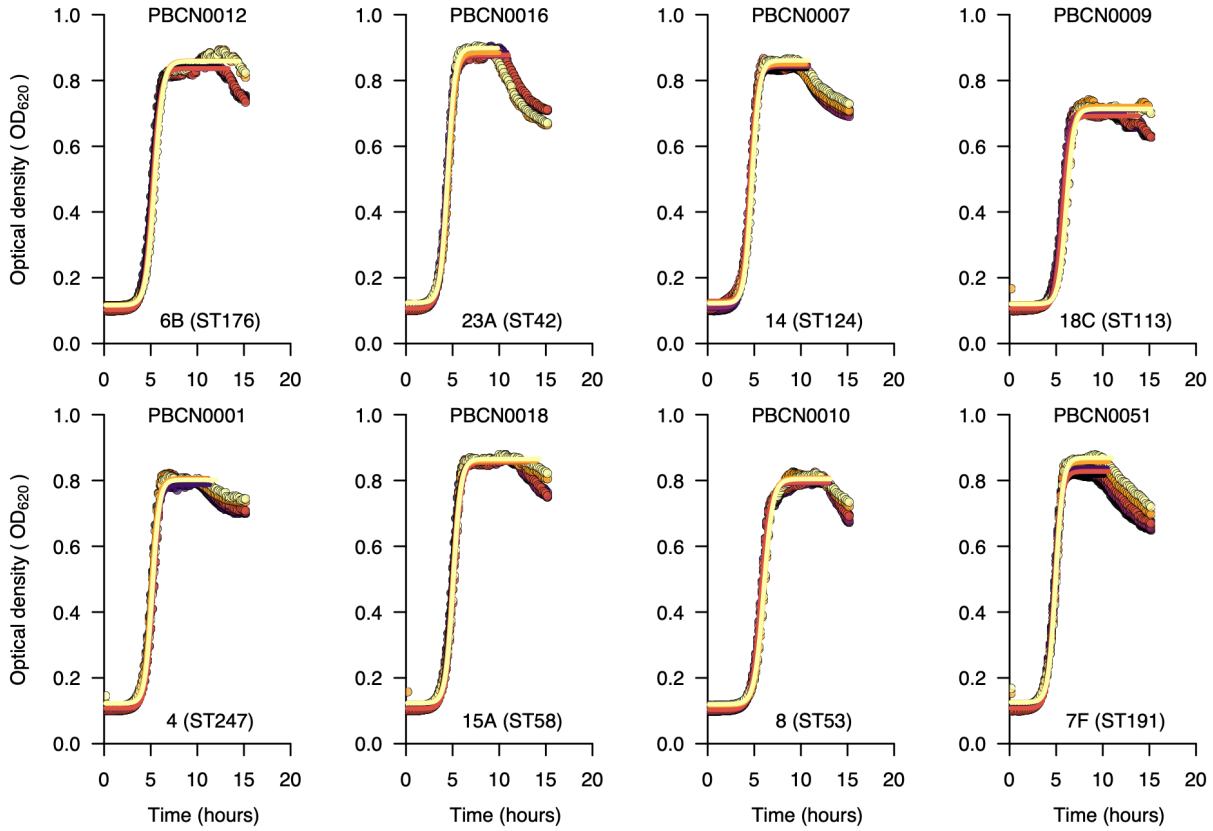


Fig. 1. Schematic overview of high-throughput quantification of *in vitro* pneumococcal growth kinetics and the genome-wide discovery of individual genomic variation influencing *in vitro* pneumococcal growth. Population genomic approaches using genome-wide association study (GWAS) offer a robust way to identify genomic variation influencing the growth dynamics of the pneumococcus. First, we collected 348 pneumococcal blood culture isolates that were obtained from 348 patients admitted to the hospital with invasive pneumococcal disease. The isolates were inoculated in a liquid medium in a highly standardised manner followed by measurement of the growth kinetics after every ten minutes for slightly over 15 hours using a microplate reader[13]. We then fitted a logistic function (see methods) to measure growth features, namely, the average growth rate (r), maximum growth density (H_{max}) defined as L_{max}, lag phase duration, and maximum growth change (ΔH), defined as the difference between the maximum

842 and minimum density (L_{max}-L_{min}). Additionally, genomic DNA was extracted from the cultured
843 isolates for whole genome sequencing using the Illumina HiSeq 4000 sequencing platform. The
844 generated sequencing data was used to reconstruct a maximum-likelihood phylogenetic tree of
845 the isolates for downstream analyses, for example, the assessment of phylogenetic signals for the
846 estimated growth features. We also performed *de novo* assembly of the sequencing reads to
847 generate contigs, which were used to map against the ATCC 700669 pneumococcal reference
848 genome (GenBank accession: NC_011900) to identify consensus single nucleotide
849 polymorphisms (SNPs). We also generated presence, and absence patterns for unitigs or variable
850 length *k*-mer sequences based on the draft assembled contigs. The SNPs were used to generate
851 the phylogeny of the isolates while the unitigs were used in a GWAS to identify variants
852 associated with the estimated pneumococcal growth features. The images were created with
853 BioRender (https://www.biorender.com/).



854

855 **Fig. 2. Growth curves for selected serotypes show the *in vitro* pneumococcal growth**
856 **dynamics and estimated growth features inferred by mathematical modelling.** The growth
857 curves above show the optical density at 620 nm monitored over 15 hours for an example dataset
858 of 20 pneumococcal isolates representing different serotypes and sequence types (STs) based on
859 the pneumococcal MLST scheme. We used the nonlinear optimisation algorithm to identify a
860 combination of parameters with the best fit to a logistic curve based on the lowest residual sum
861 of squares for the predicted value based on the model relative to the observed growth data (see
862 methods). We fitted separate curves based on a subset of points for each replicate up to the end
863 of the stationary phase or end of the experiment, whichever came first. We did this to get a better
864 fit for the logistic curve during the growth phase, as it does not allow for negative growth, which
865 happened after the stationary phase. We fitted the curves to infer the growth features for each of
866 the six replicates. The averaged estimates for each replicate were then used for the downstream
867 analyses. Additional plots showing fitted curves for all the isolates are shown in **Supplementary**
868 **Fig. 1**.

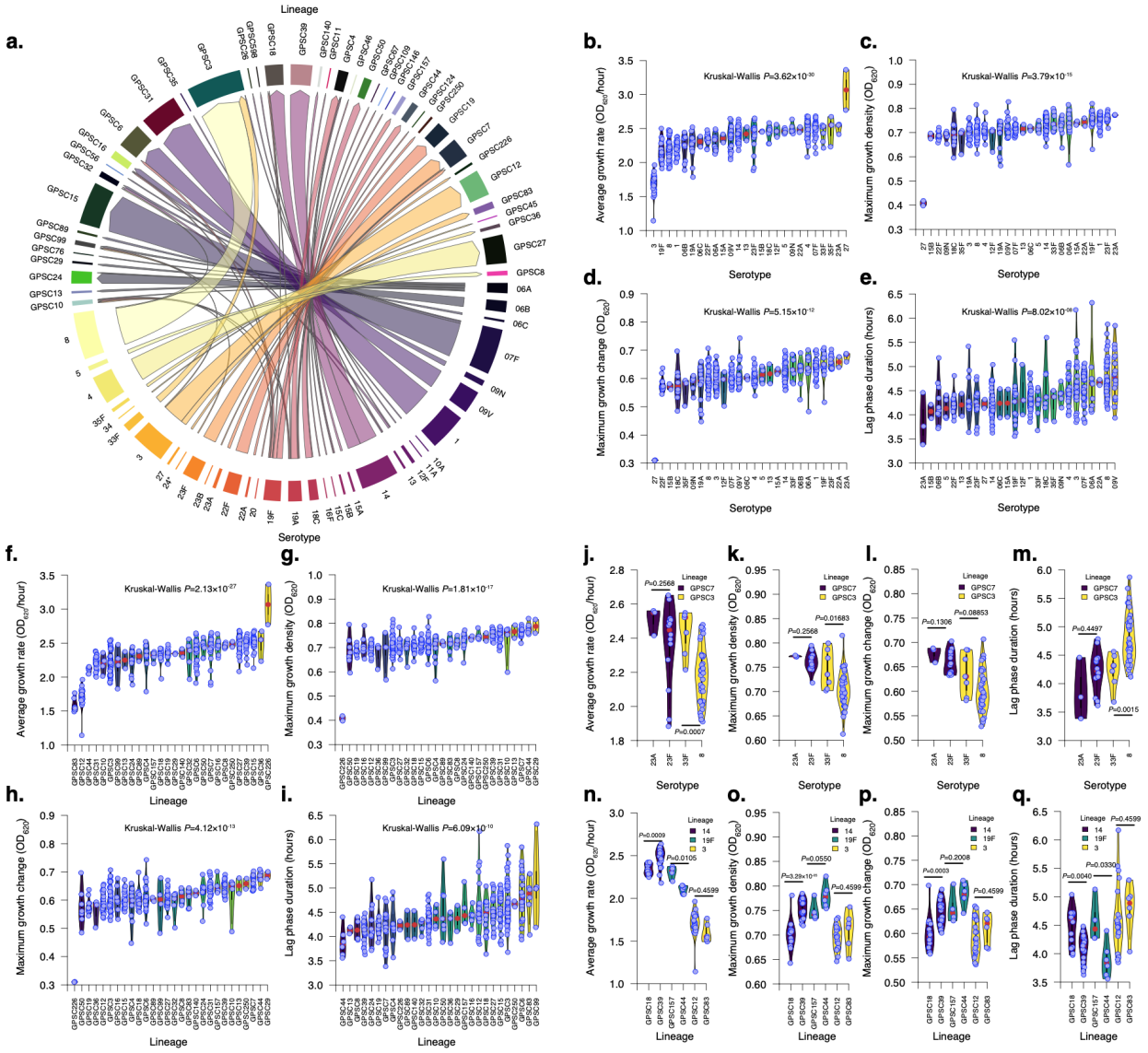


Fig. 3. The *in vitro* pneumococcal growth features show variable patterns dependent on the capsular serotype and genetic background or lineage. (a) Chord diagram showing the association between pneumococcal capsular serotypes and lineages for the Dutch isolates included in this study. The serotypes were determined using *in silico* approaches based on the sequencing data using SeroBA[85], while the lineages were defined using the PopPUNK framework[86] based on the global pneumococcal sequencing cluster (GPSC) lineage nomenclature[34]. The thickness of the connecting lines between serotypes and lineages in the chord diagram represents the percentage of isolates belonging to each serotype and lineage. Due to capsule switching, some serotypes were associated with multiple lineages, and similarly, some lineages were associated with multiple serotypes. The violin plot shows the distribution of the estimated average growth rate across different serotypes. The distribution of the estimated

881 growth features varied across serotypes with at least three isolates, as shown by the violin plots
882 for the **(b)** average growth rate, **(c)** maximum growth change, **(d)** maximum growth change
883 across serotypes, **(e)** lag phase duration growth features. Similarly, the distribution of the
884 estimated growth features varied across lineages with at least three isolates, as shown by the
885 violin plots for the **(f)** average growth rate, **(g)** maximum growth change, **(h)** maximum growth
886 density, and **(i)** lag phase duration growth features. A comparison of the growth features between
887 two serotypes belonging to the same lineages with at least three isolates is shown for the **(j)**
888 growth rate, **(k)** maximum growth, **(l)** maximum growth change, and **(m)** the lag phase duration
889 features. A comparison of the growth features between two lineages expressing the same
890 serotype containing at least three isolates are shown for the **(n)** growth rates, **(o)** maximum
891 growth, **(p)** maximum growth change, and **(q)** the lag phase features. Additional information
892 regarding the isolates and their estimated growth features or parameters are available in
893 **Supplementary Data 2**.

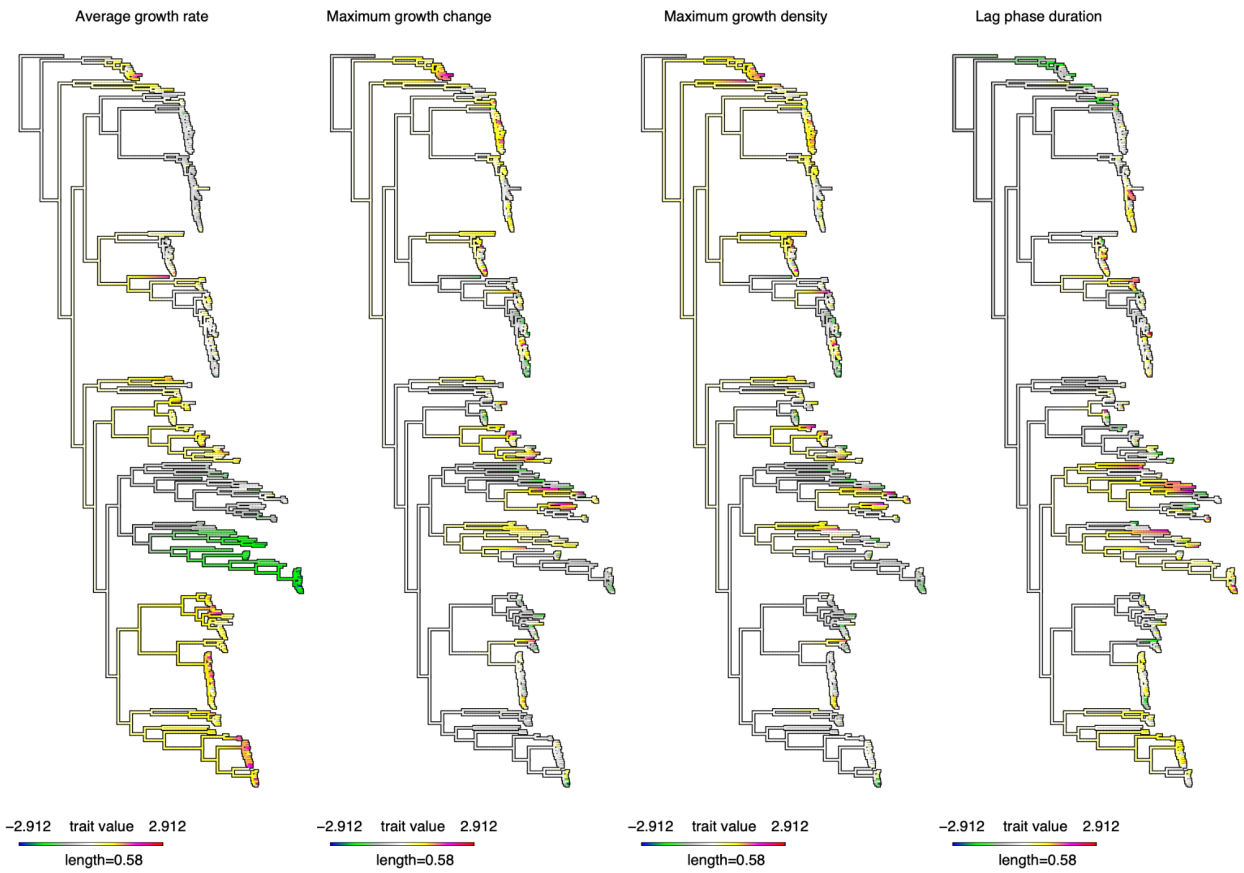


Fig. 4. Ancestral reconstruction reveals a strong correlation between the estimated *in vitro* pneumococcal growth features and the whole-genome phylogeny of the isolates. The evolution of the normalised estimated continuous growth features, namely average growth rate, maximum growth change, maximum growth density, and lag phase duration, were assessed by estimating the states of the internal nodes of the maximum likelihood phylogenetic tree of the isolates using a maximum likelihood ancestral reconstruction approach[92]. The length of the scale bar shows the number of nucleotide substitutions per site, while the colours represent the lowest and highest unnormalised growth features. The length of the scale bar shows the number of nucleotide substitutions per site, while the colours represent the lowest and highest unnormalised growth features. A phylogenetic tree showing bootstrap support values on the internal nodes of the tree is provided in **Supplementary Data 3**.

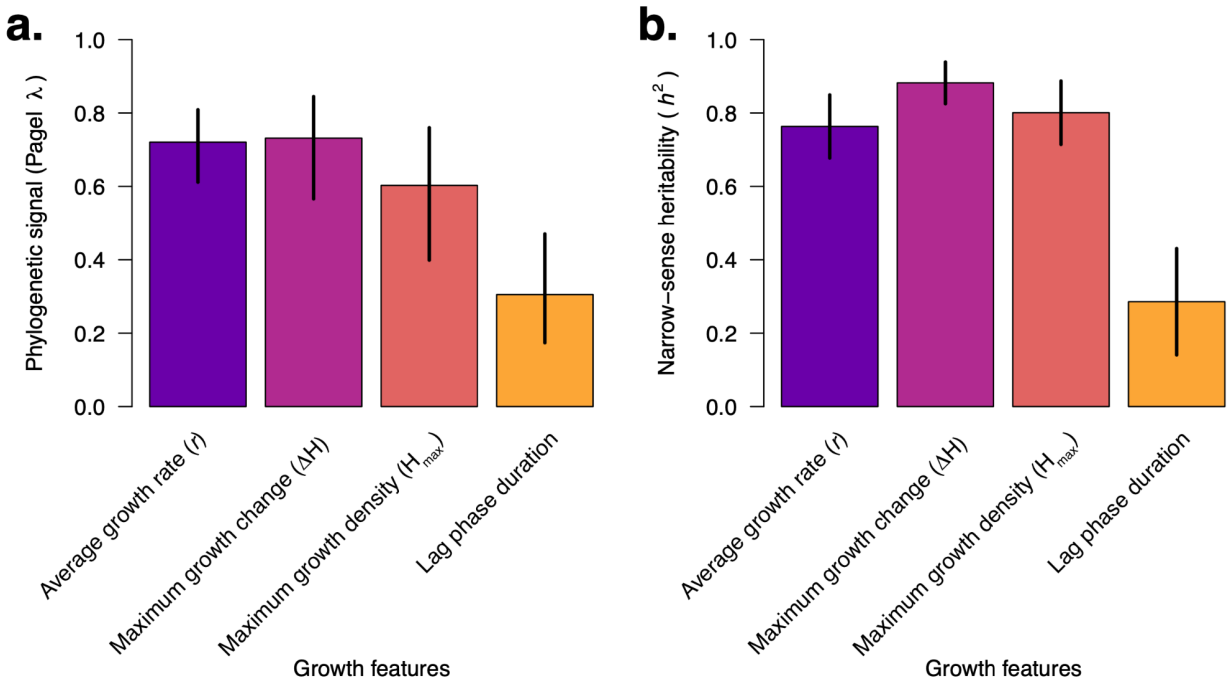
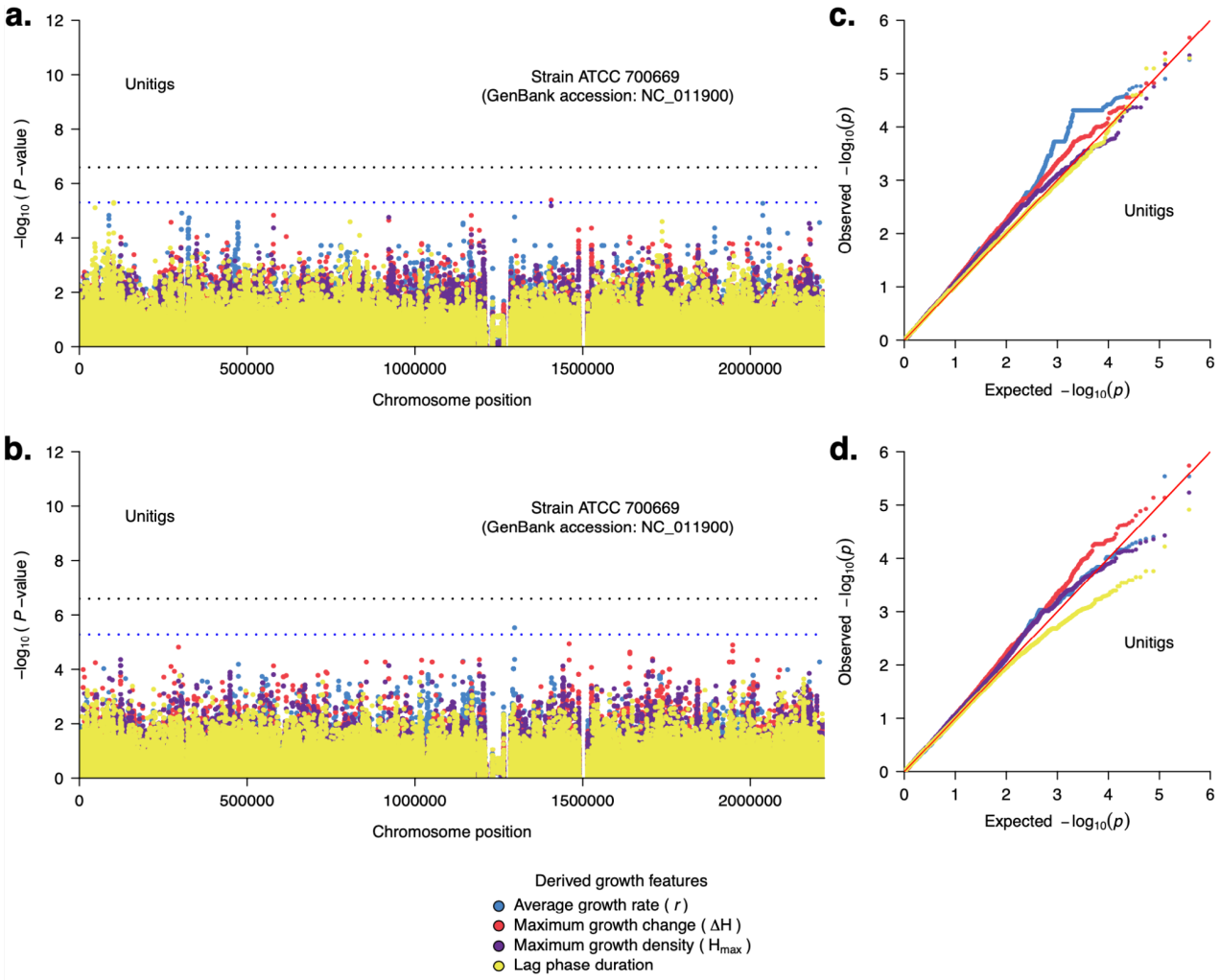


Fig. 5. The estimated *in vitro* pneumococcal growth features show high phylogenetic signals and narrow-sense heritability. (a) Bar plot showing the phylogenetic signal of the growth features using Pagel's λ statistic[38]. Pagel's λ measures the correlated evolution of the traits with the phylogenetic tree of the isolates. Pagel's λ estimates close to zero indicate no phylogenetic signal, i.e., the growth features were independent of the phylogeny, while values close to one show a strong phylogenetic signal. The average growth rate, maximum growth change, and maximum growth density showed the highest phylogenetic signal, while the lag phase duration feature showed an intermediate signal. In contrast, the maximum growth rate defined by Arends et al.[13] had the lowest phylogenetic signal, i.e., evolved more independently of the phylogeny. **(b)** Bar plot showing the narrow-sense heritability (h^2) of the pneumococcal growth features based on unitig sequences. The narrow-sense heritability corresponds to the amount of variability in the traits, i.e., growth features, explained by the pneumococcal genetic variation. The heritability is the proportion of phenotypic variation explained by the bacterial genetics (PVE) parameter using the linear mixed effects model implemented in GEMMA[96]. A heritability value close to zero implies low heritability or minimal contribution of genetics to the variability of the traits. In contrast, values close to one suggest a strong influence of genetics on growth features. There was a substantial impact of pneumococcal genetics on all the growth features, with the highest influence seen for the average growth rate, maximum growth change, and maximum growth density, with a slightly lower contribution for the lag phase duration and the maximum

growth rate measured by Arends et al.[13]. All the error bars represent the 95% confidence intervals. Additional information is provided in **Supplementary Data 3**.



928

929 **Fig. 6. Pneumococcal genetic variation associated with the *in vitro* growth features or**
930 **parameters of the isolates.** Manhattan plot showing the statistical significance of the hits from
931 the GWAS analysis based on unitigs **(a)** without adjusting for the serotype and **(b)** after adjusting
932 for the serotype. The statistical significance (*P*-value) were inferred based on the likelihood ratio
933 test using Pyseer[94] mapped against the ATCC 700669 pneumococcal reference genome
934 (GenBank accession: NC_011900[101]) to identify their location in the pneumococcal genome. The
935 dots in the Manhattan plots with different colours represent GWAS for different growth features,
936 as shown in the key. The red dotted line designates the genome-wide threshold for considering a
937 variant as significantly associated, statistically, with the growth features. The blue dotted line
938 represents the threshold for classifying variants as suggestively associated with the growth
939 features. **(c)** Quantile–quantile plots showing the relationship between the observed statistical
940 significance and the expected statistical significance for the GWAS of growth features not
941 adjusting for the serotype. **(d)** Quantile–quantile (QQ) plots showing the relationship between

942 the observed statistical significance and the expected statistical significance for the GWAS of
943 growth features after adjusting for the capsular serotype.

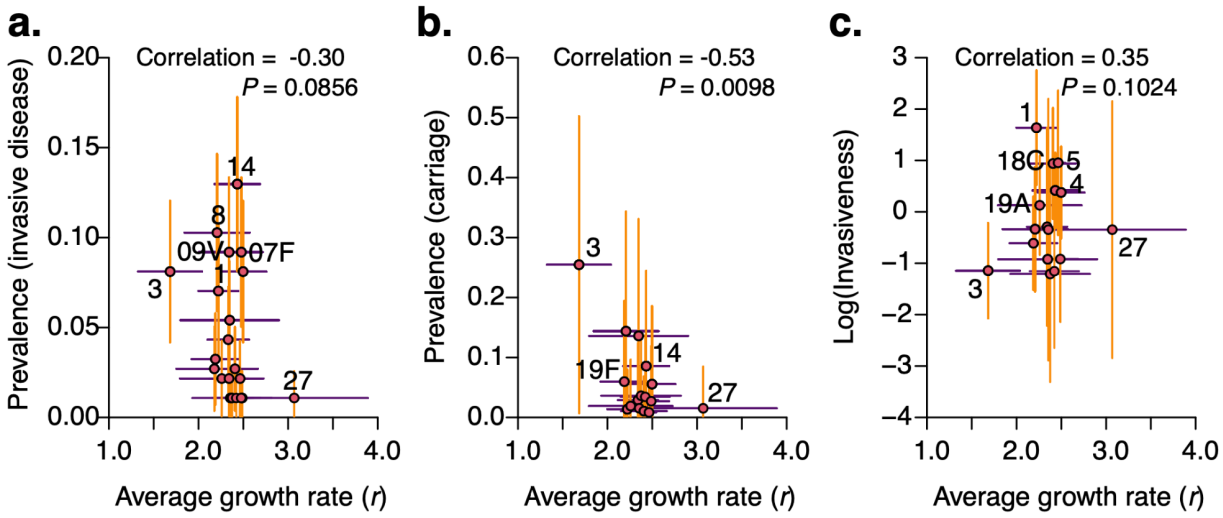


Fig. 7. Association between average *in vitro* pneumococcal growth rate with the prevalence of serotypes in invasive disease, nasopharyngeal carriage, and invasiveness. (a) Scatter plot showing the relationship between the prevalence of serotypes among invasive disease isolates collected before the introduction of PCV7 in the Netherlands and the average *in vitro* growth rate (*r*) of the isolates inferred as shown in Fig. 1. A Spearman correlation test showed no evidence of an association between the growth rate and prevalence of serotypes among the invasive disease isolates. **(b)** Scatter plot showing the relationship between the inferred prevalence of serotypes in the nasopharyngeal carriage before the introduction of PCV7 in the Netherlands and the inferred average *in vitro* growth rate. Due to the unavailability of the nasopharyngeal carriage data, we imputed the prevalence of the serotypes in carriage by dividing the prevalence in invasive disease by the average invasiveness of each serotype. The invasiveness index was calculated based on the meta-analysis of data collected before PCV7 introduction across multiple countries (see methods). A Spearman correlation test of the growth rate and prevalence of serotypes in carriage suggested a negative association between the variables. The association remained even after excluding serotype 3, which seemed like an outlier. **(c)** The relationship between the invasiveness and the average growth rates of each serotype from the pre-vaccination data. A Spearman correlation test of the growth rates and invasiveness suggested no association between the variables. All the error bars represent the 95% confidence intervals. Additional information is provided in **Supplementary Data 4**.